

Restaurant tips dataset

Q Search Sheet

Home

Insert

Draw

Page Layout

Formulas

Data

Review

View

Developer

Share

Office Update

To keep up to date with security updates, fixes and improvements, choose Check for Updates.

Check for Updates

A2

Female

sex	smoker	day	time	size	total_bill	tip
Female	No	Sun	Dinner	2	16.99	1.01
Male	No	Sun	Dinner	3	10.34	1.66
Male	No	Sun	Dinner	3	21.01	3.5
Male	No	Sun	Dinner	2	23.68	3.31
Female	No	Sun	Dinner	4	24.59	3.61
Male	No	Sun	Dinner	4	25.29	4.71
Male	No	Sun	Dinner	2	8.77	2
Male	No	Sun	Dinner	4	26.88	3.12
Male	No	Sun	Dinner	2	15.04	1.96
Male	No	Sun	Dinner	2	14.78	3.23
Male	No	Sun	Dinner	2	10.27	1.71
Female	No	Sun	Dinner	4	35.26	5
Male	No	Sun	Dinner	2	15.42	1.57
Male	No	Sun	Dinner	4	18.43	3
Female	No	Sun	Dinner	2	14.83	3.02
Male	No	Sun	Dinner	2	21.58	3.92
Female	No	Sun	Dinner	3	10.33	1.67
Male	No	Sun	Dinner	3	16.29	3.71
Female	No	Sun	Dinner	3	16.97	3.5
Male	No	Sat	Dinner	3	20.65	3.35
Male	No	Sat	Dinner	2	17.92	4.08
Female	No	Sat	Dinner	2	20.29	2.75
Female	No	Sat	Dinner	2	15.77	2.23
Male	No	Sat	Dinner	4	39.42	7.58
Male	No	Sat	Dinner	2	19.82	3.18
Male	No	Sat	Dinner	4	17.81	2.34
Male	No	Sat	Dinner	2	13.37	2
Male	No	Sat	Dinner	2	12.69	2
Male	No	Sat	Dinner	2	21.7	4.3
Female	No	Sat	Dinner	2	19.65	3
Male	No	Sat	Dinner	2	9.55	1.45
Male	No	Sat	Dinner	4	18.35	2.5

Encoded Categorical Values to Numeric Values

sex	Gender of the customer	M=0, F=1
smoker	Indicates if the customer is a smoker or not	Smoker no=0, smoker yes=1
day	Day of the restaurant visit	Thur=0, Fri=1, Sat=2, Sun=3
time	Indicates whether the tip was for lunch or dinner	Dinner=0, Lunch=1
size	Number of members dining	Already in Numbers
total_bill	Bill amount in USD	Already in Numbers
tip	Tip amount in USD	Already in Numbers

Problem Statement - We have to Predict the Restaurant Tips with given Input values with the Mathematical equation

I defined the Solution of Predictive Restaurant tips in the working sheet 2 and RMSE in the Performance Sheet

tips

Sex

Smoker

Day

Time

Size

Total_bill

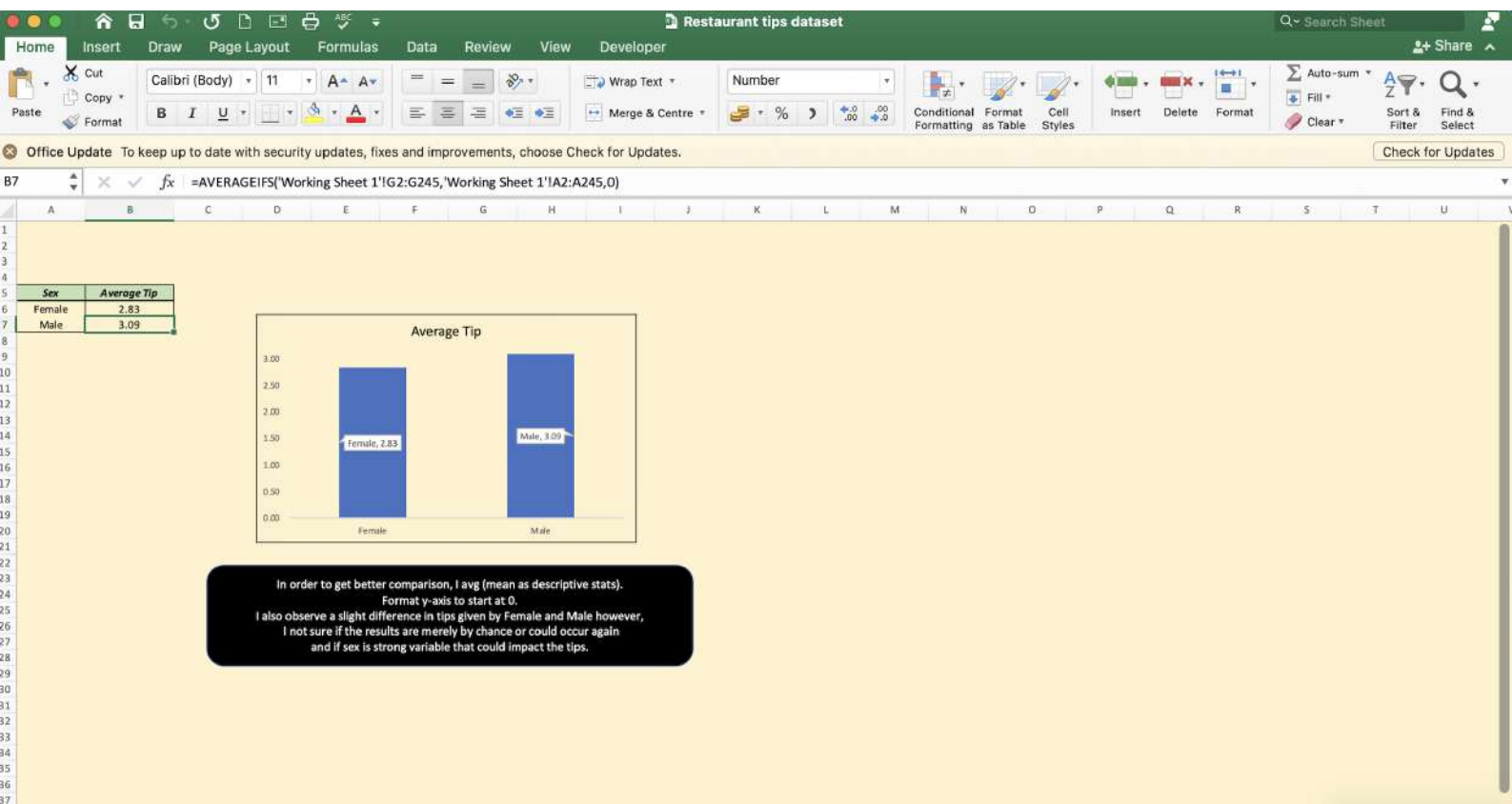
Working Sheet 1

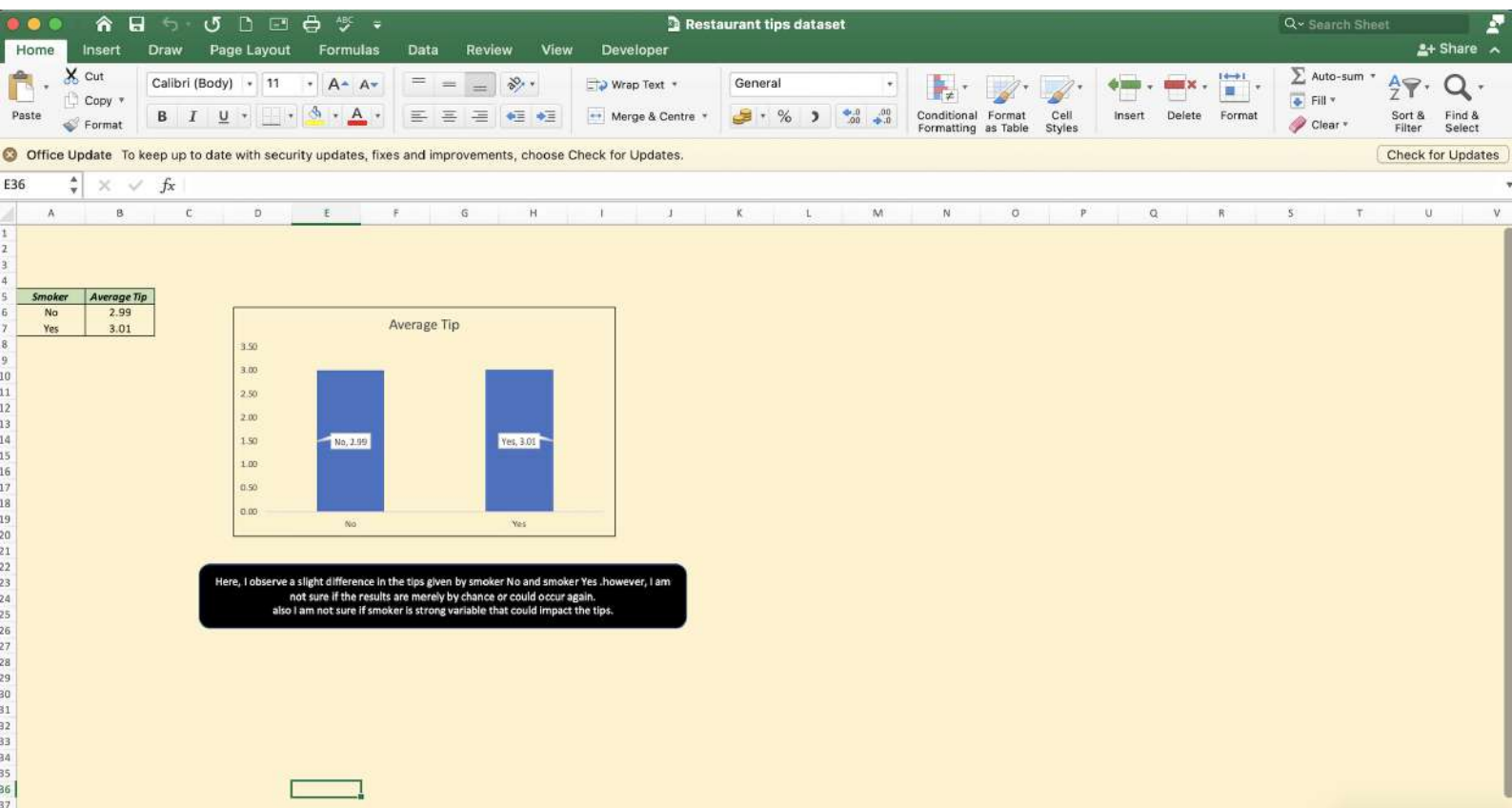
Multi Linear Regression Model

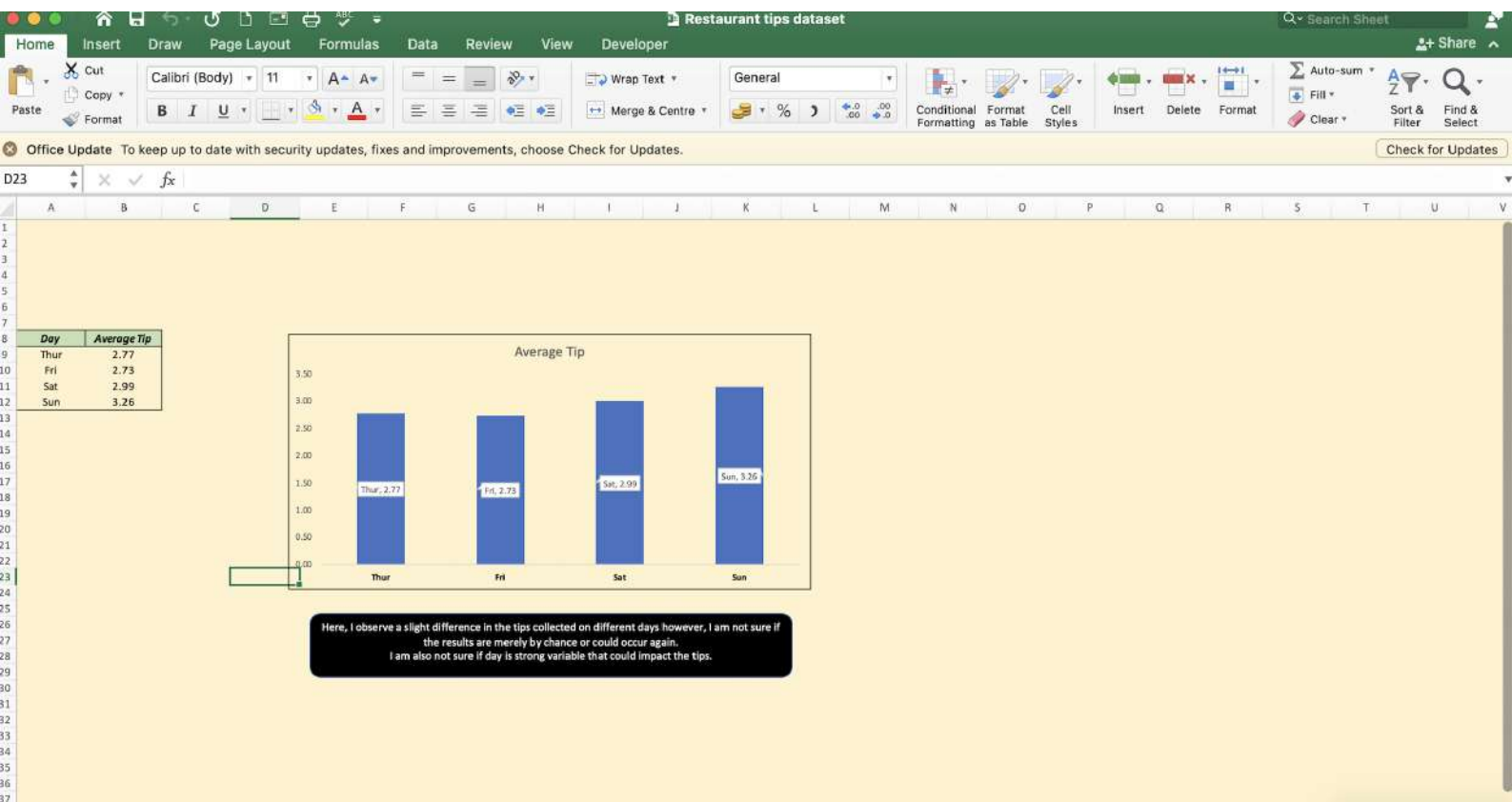
MLR 2 - Tips Predictor

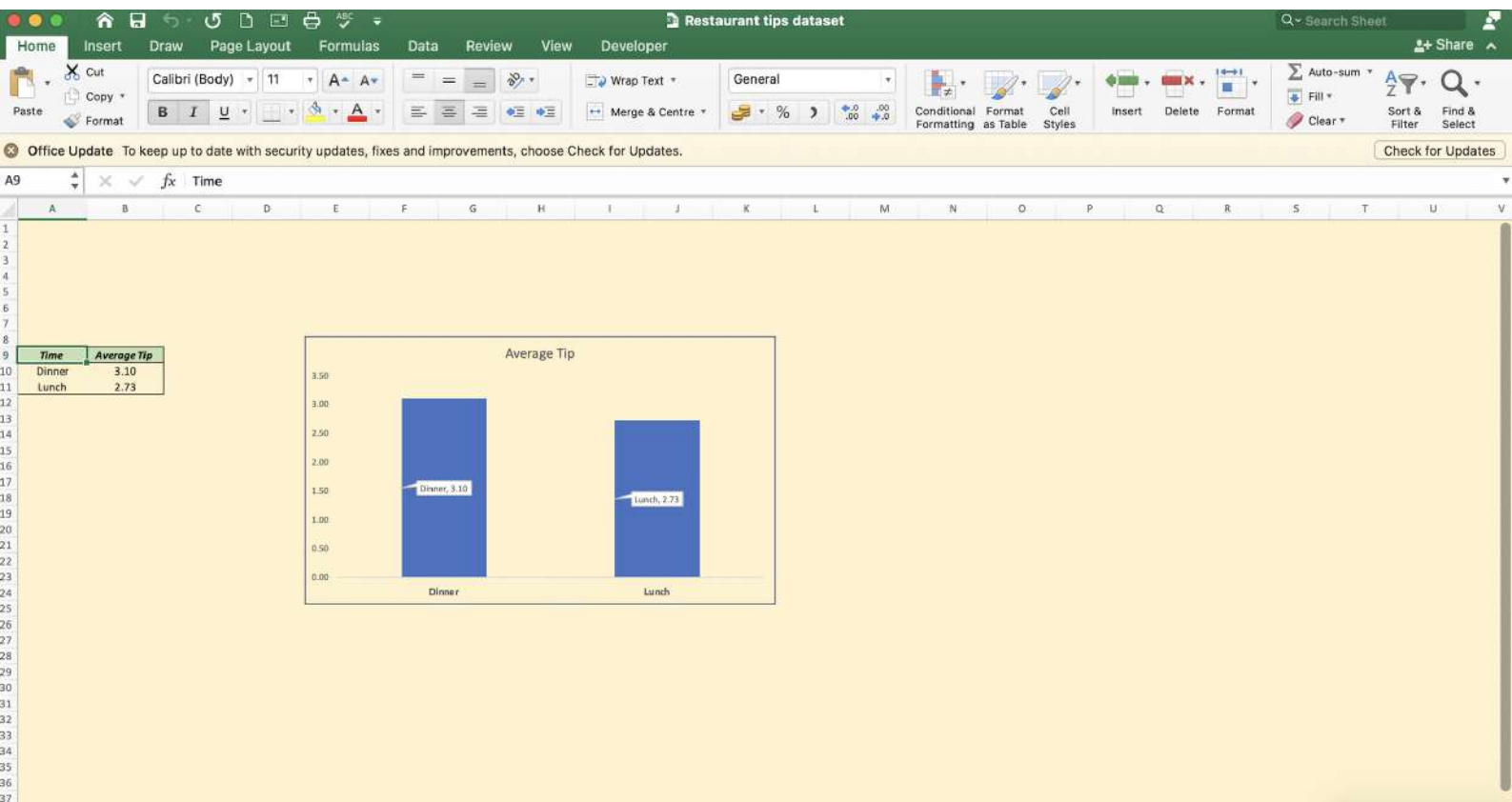
Actual vs Predictive Tips &RMSE

+









Restaurant tips dataset

Search Sheet

HomeInsertDrawPage LayoutFormulasDataReviewViewDeveloper

CutCopyPasteFormat

Calibri (Body)11A A

B I U

Wrap Text

General

Conditional FormattingFormat as TableCell Styles

InsertDeleteFormat

Auto-sumFillClear

Sort & FilterFind & Select

Office UpdateTo keep up to date with security updates, fixes and improvements, choose Check for Updates.

Check for Updates

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	Size	Total_bill	Actual Tip	Predictive Tip	Error	Absolute Error	Square Error	Abs % Error										
2	2	16.99	1.01	2.63	-1.62	1.62	2.62	1.60										
3	3	10.34	1.66	2.21	-0.55	0.55	0.30	0.33										
4	3	21.01	3.5	3.19	0.31	0.31	0.09	0.09										
5	2	23.68	3.31	3.25	0.06	0.06	0.00	0.02										
6	4	24.59	3.61	3.72	-0.11	0.11	0.01	0.03										
7	4	25.29	4.71	3.78	0.93	0.93	0.86	0.20										
8	2	8.77	2	1.87	0.13	0.13	0.02	0.07										
9	4	26.88	3.12	3.93	-0.81	0.81	0.66	0.26										
10	2	15.04	1.96	2.45	-0.49	0.49	0.24	0.25										
11	2	14.78	3.23	2.42	0.81	0.81	0.65	0.25										
12	2	10.27	1.71	2.01	-0.30	0.30	0.09	0.17										
13	4	35.26	5	4.71	0.29	0.29	0.09	0.06										
14	2	15.42	1.57	2.48	-0.91	0.91	0.83	0.58										
15	4	18.43	3	3.15	-0.15	0.15	0.02	0.05										
16	2	14.83	3.02	2.43	0.59	0.59	0.35	0.20										
17	2	21.58	3.92	3.05	0.87	0.87	0.75	0.22										
18	3	10.33	1.67	2.20	-0.53	0.53	0.29	0.32										
19	3	16.29	3.71	2.76	0.95	0.95	0.91	0.26										
20	3	16.97	3.5	2.82	0.68	0.68	0.46	0.19										
21	3	20.65	3.35	3.16	0.19	0.19	0.04	0.06										
22	2	17.92	4.08	2.72	1.36	1.36	1.86	0.33										
23	2	20.29	2.75	2.94	-0.19	0.19	0.03	0.07										
24	2	15.77	2.23	2.52	-0.29	0.29	0.08	0.13										
25	4	39.42	7.58	5.09	2.49	2.49	6.18	0.33										
26	2	19.82	3.18	2.89	0.29	0.29	0.08	0.09										
27	4	17.81	2.34	3.09	-0.75	0.75	0.56	0.32										
28	2	13.37	2	2.29	-0.29	0.29	0.09	0.15										
29	2	12.69	2	2.23	-0.23	0.23	0.05	0.12										
30	2	21.7	4.3	3.07	1.23	1.23	1.52	0.29										
31	2	19.65	3	2.88	0.12	0.12	0.02	0.04										
32	2	9.55	1.45	1.94	-0.49	0.49	0.24	0.34										
33	4	18.35	2.5	3.14	-0.64	0.64	0.41	0.26										
34	2	15.06	3	2.45	0.55	0.55	0.30	0.18										
35	4	20.69	2.45	3.36	-0.91	0.91	0.82	0.37										
36	2	17.78	3.27	2.70	0.57	0.57	0.32	0.17										
37	3	24.06	3.6	3.48	0.12	0.12	0.02	0.03										

Y=Constant +B1*(X1)+B2*(X2)+....BnXn

Constant = Intercept

X=Coefficients

B=Actual independent variable values

	Intercept (B)	X1(Size)	X2(Total_bill)
Coefficients	0.668944741	0.192597794	0.092713337

Sum of Square Error (SSE)

Mean of Square Error (MSE)

Root Mean Square Error (RMSE)

247.55

1.01

0.51

So, Based on my analysis, I can conclude that based on size and Total bill are the best Independent variables through which I can get the Best Predictive Tip for the Restaurant and also I calculated the error which is basically a difference of the change in tips.

tipsSexSmokerDayTimeSizeTotal_billWorking Sheet 1Multi Linear Regression ModelMLR 2 - Tips PredictorActual vs Predictive Tips &RMSE+

Restaurant tips dataset

Search Sheet

Share

HomeInsertDrawPage LayoutFormulasDataReviewViewDeveloper

Paste

Cut

Copy

Format

Calibri (Body)

11

A

A

B

I

U

A

Wrap Text

General

Conditional Formatting

Format as Table

Cell Styles

Insert

Delete

Format

Auto-sum

Fill

Clear

Sort & Filter

Find & Select

Office Update

To keep up to date with security updates, fixes and improvements, choose Check for Updates.

Check for Updates

L16

From the all Independent Variables data, I can conclude that size and total_bill might be good indicators to predict tips.

I need further analysis to be sure and find how exactly it will impact the tips.

For variables -sex,smoker,day,time I compared mean of categories to the tips and I cannot draw any conclusions which could point that these variables might impact tips.

Further analysis is required for the same.

I intent to do the analysis using Multi linear Regression using the data analysis toolpak in excel.

Dependent variable is tips.

Independent variables are sex,smoker,day,time ,size and total_bill

I converted the categorical variables into numeric ones in this sheet.

Restaurant tips dataset

Search Sheet

HomeInsertDrawPage LayoutFormulasDataReviewViewDeveloper

Cut

Copy

Paste

Calibri (Body)

11

A

B

I

U

Wrap Text

Merge & Centre

General

Conditional Formatting

Format as Table

Cell Styles

Insert

Delete

Format

Auto-sum

Fill

Clear

Sort & Filter

Find & Select

Office UpdateTo keep up to date with security updates, fixes and improvements, choose Check for Updates.

Check for Updates

135

fx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1	SUMMARY OUTPUT		This Statistics is based on all Independent variables																
2																			
3	Regression Statistics																		
4	Multiple R	0.684980787	68%																
5	R Square	0.469198679	46%																
6	Adjusted R Square	0.455760671	45%																
7	Standard Error	1.020745565																	
8	Observations	244																	
9																			
10	ANOVA																		
11		df	SS	MS	F	Significance F													
12	Regression	6	218.2770796	36.37951327	34.91579067	4.09922E-30													
13	Residual	237	246.9353974	1.041921508															
14	Total	243	465.212477																
15																			
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%												
17	Intercept	0.576880629	0.355005994	1.624988419	0.105494101	-0.122489703	1.27625096												
18	sex	0.034644964	0.141081963	0.245566218	0.806230561	-0.24328989	0.312579818												
19	smoker	-0.075663089	0.140198277	-0.539686293	0.589920088	-0.351857061	0.200530884												
20	day	0.05273982	0.120334639	0.4382763	0.661585219	-0.184322308	0.289801948												
21	time	0.112477769	0.307526134	0.365750277	0.714877667	-0.493356099	0.718311636												
22	size	0.174819618	0.089187194	1.960142606	0.051150876	-0.000881295	0.350520531												
23	total_bill	0.094325088	0.009538173	9.889219168	1.57818E-19	0.075534657	0.113115518												
24																			
25																			
26																			
27	p value less than 5%--> statistical significant relationship.																		
28	eg: p value= 0.919003636 then there is 91% chance that it can be																		
29	explained by randomness.																		
30	Exclude columns with p values greater than 0.05																		
31																			
32																			
33																			
34																			
35																			

Restaurant tips dataset

Office Update To keep up to date with security updates, fixes and improvements, choose Check for Updates.

C21

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	SUMMARY OUTPUT		This Statistics are based on size & total_bill												
2															
3	Regression Statistics														
4	Multiple R	0.684009729	68%												
5	R Square	0.467869309	46%												
6	Adjusted R Square	0.463453286	46%												
7	Standard Error	1.013505967													
8	Observations	244													
9															
10	ANOVA														
11		df	SS	MS	F	Significance F									
12	Regression	2	217.6586401	108.82932	105.9481301	9.66509E-34									
13	Residual	241	247.553837	1.027194344											
14	Total	243	465.212477												
15															
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%								
17	Intercept	0.688944741	0.193609331	3.455126549	0.000649806	0.287562197	1.050327285								
18	size	0.192597794	0.085314557	2.25750214	0.024872446	0.024540385	0.360655204								
19	total_bill	0.092713337	0.009114682	10.17186688	1.88092E-20	0.074758723	0.110667951								
20															
21															
22															
23															
24															
25															
26															
27															
28															
29															
30															
31															
32															

